

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/100736/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Hanson, W. Miachel, Chen, Zhe, Jackson, Laurie K., Attaf, Meriem, Sewell, Andrew K. ORCID: <https://orcid.org/0000-0003-3194-3135>, Heemstra, Jennifer M. and Phillips, John D. 2016. Reversible oligonucleotide chain blocking enables bead capture and amplification of T-Cell receptor alpha and beta chain mRNAs. Journal of the American Chemical Society 138 (35) , pp. 11073-11076. 10.1021/jacs.6b04465 file

Publishers page: <http://dx.doi.org/10.1021/jacs.6b04465>  
<<http://dx.doi.org/10.1021/jacs.6b04465>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





# Reversible Oligonucleotide Chain Blocking Enables Bead Capture and Amplification of T-Cell Receptor $\alpha$ and $\beta$ Chain mRNAs

W. Miachel Hanson,<sup>†,⊥</sup> Zhe Chen,<sup>‡,⊥</sup> Laurie K. Jackson,<sup>§</sup> Meriem Attaf,<sup>||</sup> Andrew K. Sewell,<sup>||</sup> Jennifer M. Heemstra,<sup>\*,‡</sup> and John D. Phillips<sup>\*,§</sup>

<sup>†</sup>Health Science Cores, <sup>‡</sup>Department of Chemistry and the Center for Cell and Genome Science, <sup>§</sup>Department of Internal Medicine, University of Utah, Salt Lake City, Utah 84112, United States

<sup>||</sup>Division of Infection & Immunity, Cardiff University School of Medicine, Cardiff, United Kingdom

## S Supporting Information

**ABSTRACT:** Next-generation sequencing (NGS) has proven to be an exceptionally powerful tool for studying genetic variation and differences in gene expression profiles between cell populations. However, these population-wide studies are limited by their inability to detect variation between individual cells within a population, inspiring the development of single-cell techniques such as Drop-seq, which add a unique barcode to the mRNA from each cell prior to sequencing. Current Drop-seq technology enables capture, amplification, and barcoding of the entire mRNA transcriptome of individual cells. NGS can then be used to sequence the 3'-end of each message to build a cell-specific transcriptional landscape. However, current technology does not allow high-throughput capture of information distant from the mRNA poly-A tail. Thus, gene profiling would have much greater utility if beads could be generated having multiple transcript-specific capture sequences. Here we report the use of a reversible chain blocking group to enable synthesis of DNA barcoded beads having capture sequences for the constant domains of the T-cell receptor  $\alpha$  and  $\beta$  chain mRNAs. We demonstrate that these beads can be used to capture and pair TCR $\alpha$  and TCR $\beta$  sequences from total T-cell RNA, enabling reverse transcription and PCR amplification of these sequences. This is the first example of capture beads having more than one capture sequence, and we envision that this technology will be of high utility for applications such as pairing the antigen receptor chains that give rise to autoimmune diseases or measuring the ratios of mRNA splice variants in cancer stem cells.

The widespread availability of next-generation sequencing (NGS) instruments has enabled researchers around the world to examine organisms at unprecedented levels of detail at low cost, resulting in a rapid escalation of NGS-based studies.<sup>1</sup> Current technologies generally require large amounts of template material, which is produced from samples containing millions of cells. While data generated using this approach give a view of the population-wide characteristics of these cells, it provides no mechanism for linking information to individual cells. As a result, it is not possible to directly examine the distribution of variation across large populations of cells within a sample. This important limitation means that variation that is

present at a low frequency within a population cannot be easily discriminated from sequencing noise. Furthermore, population-based methods are incapable of directly examining how the relative frequency of variation within a population of cells fluctuates over time. These deficiencies of NGS technology pose a particularly large challenge for research examining cancer stem cells, which occur at a low frequency in tumors, but are thought to be central to tumor survival and treatment escape.<sup>2</sup>

Recently, high-throughput fluidics-based systems such as Drop-seq have been reported in which individual cells are incubated with an mRNA capture bead in nanoliter droplets, enabling downstream sequencing of mRNA on a single-cell level.<sup>3</sup> This method has allowed significant progress in circumventing the limitations of population-wide studies, but it remains incompatible with the preferred short read length sequencing platforms for many sequences where critical information is distant from the 3'-end of the message. Additionally, the current technology only enables capture of a single mRNA sequence (usually the poly-A tail of mRNA). The ability to use two or more capture sequences would be of significant value, for example in measuring the ratio of differentially spliced messages in individual cancer stem cells, or obtaining the paired sequences of  $\alpha$  and  $\beta$  antigen receptor chains.

The antigen receptors on individual lymphocytes control their specificity and are critical components in disease outcome.<sup>4</sup> These receptors are created by somatic gene rearrangements at two loci on different chromosomes to produce the antibody-like  $\alpha\beta$  T-cell receptor (TCR) or  $\gamma\delta$  TCR. The  $\alpha\beta$  TCR is of particular interest, as this receptor is believed to orchestrate the adaptive immune response. The critical determinant of antigen specificity is the rearrangement of variable (V), diversity (D), and joining (J) segments that make up the complementarity determining regions (CDRs), and each T-cell clone has a unique set of CDR sequences.<sup>5</sup> The antigen receptor V(D)J sequence for each chain lies adjacent to a locus-specific constant domain, enabling capture from the total cellular RNA using primers specific for this sequence.<sup>6</sup> Subsequent primer extension and PCR amplification can then allow complete chain sequence assembly from short NGS reads. Obtaining the paired  $\alpha$  and  $\beta$  chain antigen receptor

Received: April 30, 2016

Published: August 1, 2016



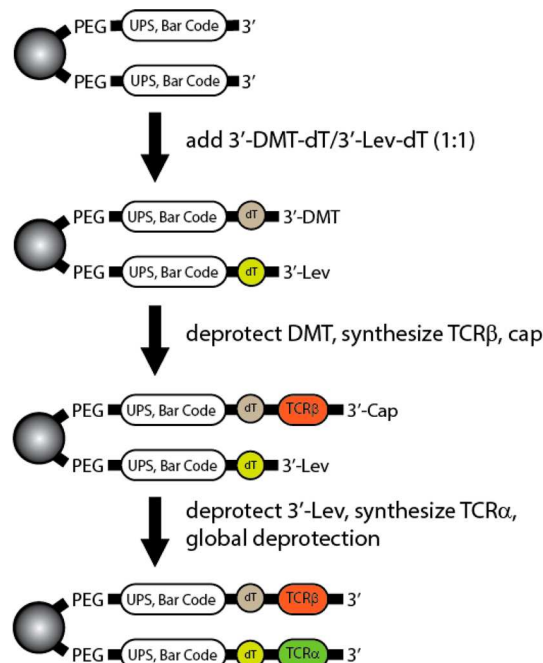
sequences for all of the lymphocytes in a clinical sample, combined with their abundance and activation status, would allow powerful systems level examination of human pathologies and disease, and there is much interest in this area.<sup>7</sup> Unfortunately, current single-cell analyses of this type are extremely laborious, constraining throughput to hundreds of cells, and therefore these studies only scratch the surface of this valuable trove of information.

The ability to obtain paired sequence information for antigen receptor chains within large clinical samples using methodologies such as Drop-seq would be a game changing technology. However, this requires the construction of capture beads having two different primer sequences attached to bead-specific unique barcodes that allow chain reconstruction from short read-length NGS platforms. Here we overcome this challenge using a novel phosphoramidite monomer that reversibly blocks chain extension to enable divergent synthesis of two primer sequences on each barcoded bead. We demonstrate that two different capture sequences can be built onto a single bead, enabling specific capture and amplification of mRNAs encoding the  $\alpha$  and  $\beta$  chains of the  $\alpha\beta$  TCR.

In principle, capture beads having two different primer sequences could be generated by chemical attachment of the primers to beads. However, a critical component of single-cell analysis techniques is the use of unique barcodes on each bead, located upstream from the primer sequence. This enables the  $>10^8$  reads acquired from NGS to be regrouped into populations arising from each individual cell.<sup>3</sup> The barcoded beads are generated using split-pool combinatorial synthesis, and thus the primer sequences must be synthesized directly onto the barcode sequence, such that the entire sequence is amplified and read in NGS. An additional consideration for the capture beads is that the DNA must be synthesized in the “reverse” 5′–3′ direction, as this provides the 3′-terminus that is needed for enzymatic primer extension after mRNA capture.

We envisioned the synthetic strategy shown in Figure 1, in which split-pool synthesis would be used to generate barcoded beads, and these would then be reacted with a mixture of 5′ phosphoramidite monomers having two different protecting groups at the 3′-position. This would enable selective deprotection of half of the oligonucleotides and synthesis of the first primer sequence, followed by a second deprotection and primer synthesis. The first monomer could have the typical dimethoxytrityl (DMT) protecting group, which is removed under mildly acidic conditions. But, the second monomer requires a protecting group having the following characteristics: (1) stable to DMT deprotection conditions; (2) stable to conditions used for oligonucleotide synthesis; (3) deprotected under conditions compatible with DNA oligonucleotides. A search of the literature revealed no examples of 3′-protected phosphoramidite monomers that would meet these requirements. However, our interest was piqued by previous reports using a 2′-levulinyl protecting group for the synthesis of branched oligonucleotides.<sup>8</sup> This protecting group appeared to meet the requirements outlined above, as it had been shown to be stable to the conditions used for RNA synthesis and is deprotected using aqueous hydrazine, which is nondamaging to oligonucleotides and orthogonal to the acidic DMT deprotection conditions.

As shown in Figure 2a, we synthesized 3′-OLev thymine phosphoramidite **2** from commercially available **1** using standard protocols. We then carried out a synthesis of short polyT sequences to test the compatibility of **2** with our

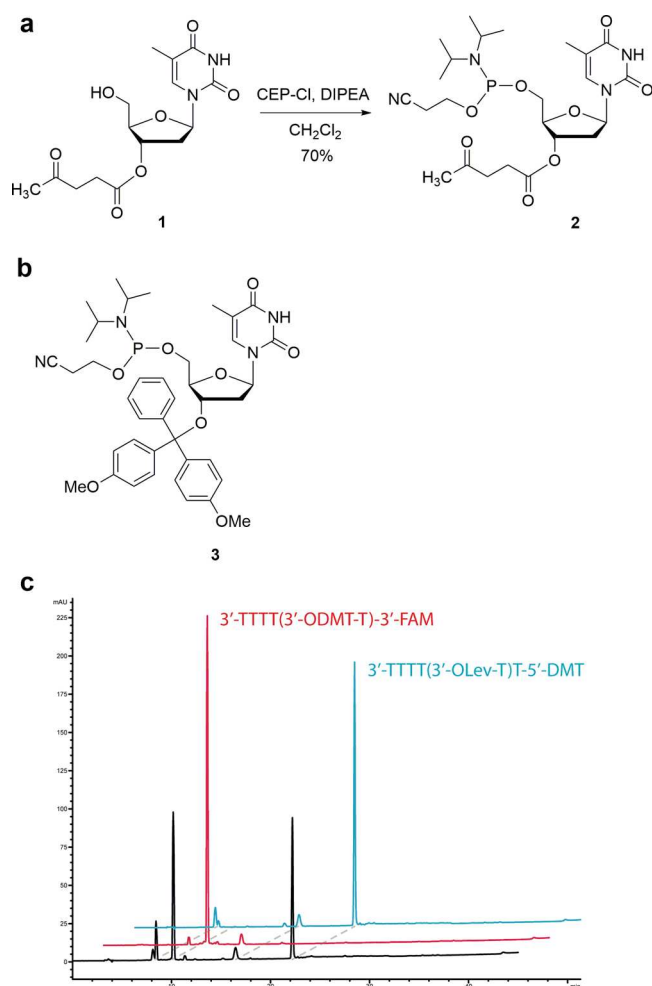


**Figure 1.** Beads can be synthesized having a unique barcode, and then a 1:1 mixture of 3′-DMT and 3′-Lev phosphoramidites are added to enable stepwise synthesis of two different oligonucleotides on each bead.

divergent oligonucleotide synthesis protocol. To mimic synthesis of the barcode region, we performed four thymine monomer coupling steps on all beads. We then added a 1:1 mixture of 3′-ODMT monomer **3** and 3′-OLev monomer **2**. The DMT group was then removed, and the deprotected oligonucleotide chains coupled with a fluorescein phosphoramidite. Next, the Lev group was removed using aqueous hydrazine, and the resulting deprotected oligonucleotide chains were coupled with an additional thymine monomer. Both oligonucleotide chains were then cleaved from the beads and analyzed by HPLC. As shown in Figure 2b, the ratio between T<sub>5</sub>-FAM DNA (arising from addition of **3**) and T<sub>6</sub> DNA (arising from addition of **2**) is very close to 1:1, demonstrating that the two different monomers are added to a growing oligonucleotide chain with nearly equal efficiency. Additionally, the high purity of the DNA products obtained indicates that the Lev protecting group is not removed during DMT removal or the subsequent coupling step and that deprotection of the Lev group does not damage the DNA molecules.

Having established that Lev can function as a DNA-compatible orthogonal protecting group, we next sought to generate dual-primer beads for capture and amplification of the mRNAs encoding the  $\alpha$  and  $\beta$  chains of the T-cell antigen receptor. Using the same synthetic protocol as described above for the test beads, we synthesized beads capable of targeting both the T-cell receptor alpha chain (TCR $\alpha$ ) and beta chain (TCR $\beta$ ) mRNAs. Specifically, the capture sequences attached to the beads are complementary to the TCR $\alpha$  and TCR $\beta$  constant regions (TRAC and TRBC). To test the ability of the capture beads to pull down TCR mRNA, the beads were mixed with total RNA isolated from a known T cell clone (clone GDB4). Unbound RNA was removed by washing, and the bead-bound RNA was subjected to a reverse transcription reaction in which the DNA oligonucleotides attached to the beads served as primers for the synthesis of bead-bound cDNA.

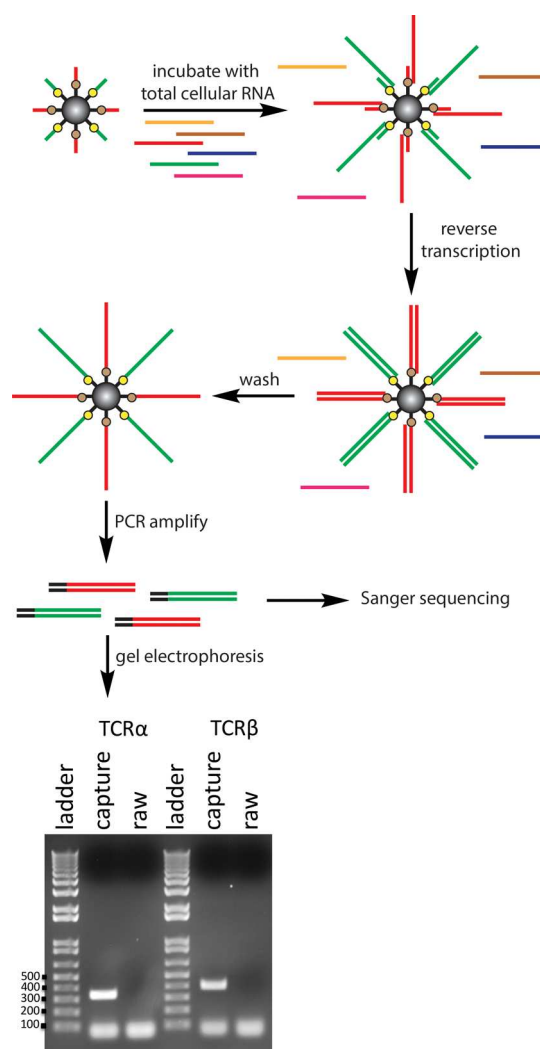




**Figure 2.** (a) Synthesis of Lev-protected monomer 2. (b) Chemical structure of DMT-protected monomer 3. (c) HPLC traces of test sequences synthesized in tandem on the same beads. Black trace shows product cleaved from beads. Blue and red traces show individually synthesized products arising from coupling of **2** and **3**, respectively.

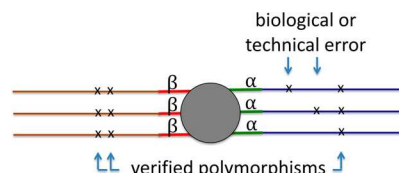
PCR amplification of the cDNA was then performed using primers complementary to the known sequence of the TCR  $\alpha$  and  $\beta$  chain transcripts in the GDB4 clone. As a control, all of the above steps were also carried out using “raw” beads, on which no DNA had been synthesized. **Figure 3** shows an agarose gel image of the PCR product from each reaction. As anticipated, we observe amplification of the TCR $\alpha$  and TCR $\beta$  sequences when using the capture beads, but not with the raw beads. The bands corresponding to the PCR products were excised from the gel and subjected to Sanger sequencing. The sequencing results shown in **Figure S1** are consistent with the known TCR $\alpha$  and TCR $\beta$  sequences in the GDB4 clone, confirming the ability of the capture beads to pull down TCR mRNA.

In conclusion, we have demonstrated that beads containing unique oligonucleotide barcodes followed by two independent oligonucleotide capture sequences can be produced. These beads can be used to capture two different specific target sequences, which has high utility in cases such as the T-cell antigen receptor, where receptor chain pairing is critical for determination of antigen specificity, but is highly variable at the single cell level. Importantly, each bead has approximately 5000 copies of each capture sequence, which provides redundancy in



**Figure 3.** RNA from lysed T cells was incubated with capture beads having sequences complementary to the constant region of TCR $\alpha$  and TCR $\beta$ . Captured RNA was reverse transcribed and PCR amplified, and then it was resolved using gel electrophoresis and visualized by ethidium bromide staining. PCR reactions from the capture beads show amplification of TCR $\alpha$  and TCR $\beta$  sequences. As a control, all of the above steps were performed using raw beads (no oligos attached) and show no PCR product. The size (in base pairs) of the smallest five bands in the DNA ladder are indicated to the left of the first gel image.

the sequencing process to discriminate real polymorphisms from transcriptional or sequencing errors (**Figure 4**). The key to achieving synthesis of the two different oligonucleotide sequences on a single bead is use of a 3'-Lev-protected phosphoramidite, which is stable to the conditions used for



**Figure 4.** Multiple copies of  $\alpha$  and  $\beta$  chain mRNAs can be captured from each cell by individual beads. Alignments of redundant sequencing reads from each bead allow discrimination between true polymorphisms and sequencing or transcriptional errors.

DMT deprotection and oligonucleotide synthesis, and can be subsequently removed without damaging the newly formed oligonucleotides. While the current study only utilized a single differential protection step, we envision that the divergent DMT/Lev protection step could be used iteratively to generate beads having larger numbers of unique DNA capture sequences. This synthesis scheme may also be used to produce capture reagents that can be used as biological probes when linked to sequences that include molecular dyes that can be specifically visualized or quenched.

The immense diversity of the adaptive immune system relies on production of three antigen receptor types, on both B-cells and T-cells, that are synthesized from transcripts that have undergone germline rearrangements. Identifying these germline rearrangements is only possible through direct sequencing of the corresponding DNA or mRNA. As our understanding of the complexity of the adaptive immune response continues to expand, abhorrent immune responses have been identified as the basis for several diseases.<sup>9</sup> These diseases can arise from a single clone, and thus the ability to obtain paired sequence information for antigen receptor  $\alpha$  and  $\beta$  chains at the single cell level could be transformative in identifying these disease causing autoimmune clones. The research reported here provides the tools necessary to undertake these experiments and demonstrates the feasibility of this approach by capturing, amplifying, and sequencing the  $\alpha$  and  $\beta$  chains for a single T-cell receptor clone. Future studies will be aimed at utilizing our capture beads for single cell sequencing of large populations of T-cell antigen receptors.

## ■ ASSOCIATED CONTENT

### ● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b04465.

Synthetic protocols and characterization data for **2**, oligonucleotide synthesis protocols and characterization, experimental details for capture and amplification of TCR mRNAs (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*john.phillips@hsc.utah.edu

\*heemstra@chem.utah.edu

### Author Contributions

<sup>†</sup>W.M.H. and Z.C. contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the HSC DNA/Peptide Core facility at the University of Utah, the National Institutes of Health (DK020503 and DK090257 supported J.D.P. and L.K.J.), and the National Science Foundation (CHE 1308364 and CHE 1608561 to J.M.H.). A.K.S. is a Wellcome Trust Senior Investigator.

## ■ REFERENCES

- (1) Schuster, S. C. *Nat. Methods* **2007**, *5*, 16–18.
- (2) Jordan, C. T.; Guzman, M. L.; Noble, M. N. *Engl. J. Med.* **2006**, *355*, 1253–1261.
- (3) Macosko, E. Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A. R.; Kamitaki, N.; Martersteck, E.

M.; Trombetta, J. J.; Weitz, D. A.; Sanes, J. R.; Shalek, A. K.; Regev, A.; McCarroll, S. A. *Cell* **2015**, *161*, 1202–1214.

(4) Castro, C. D.; Luoma, A. M.; Adams, E. J. *Immunol. Rev.* **2015**, *267*, 30–55.

(5) Margulies, D. H. *Nat. Immunol.* **2001**, *2*, 669–670.

(6) Hou, X. L.; Wang, L.; Ding, Y. L.; Xie, Q.; Diao, H. Y. *Genes Immun.* **2016**, *17*, 153–164.

(7) Attaf, M.; Huseby, E.; Sewell, A. K. *Cell. Mol. Immunol.* **2015**, *12*, 391.

(8) Katolik, A.; Johnsson, R.; Montemayor, E.; Lackey, J. G.; Hart, P. J.; Damha, M. J. *J. Org. Chem.* **2014**, *79*, 963–975.

(9) Simon, A. K.; Hollander, G. A.; McMichael, A. *Proc. R. Soc. London, Ser. B* **2015**, *282*, 20143085.